

Reutilización y adaptación de la ontología GUM al castellano

M^a del Socorro Bernardos Galindo
Laboratorio de Inteligencia Artificial
sgalindo@delicias.dia.fi.upm.es

Guadalupe Aguado de Cea
Departamento de Lingüística Aplicada
lupe@fi.upm.es

Facultad de Informática, Campus de Montegancedo s/n
28660 Boadilla del Monte (Madrid)

Resumen: En un proyecto de generación de lenguaje natural (GLN) la fuente de información del dominio no contiene normalmente los datos lingüísticos que el sistema necesita para poder generar los textos. En esta comunicación se presenta la solución adoptada en un proyecto de generación en castellano. En él se optó por utilizar una ontología que proporcionara ese conocimiento lingüístico y en la cual se clasificara la información del dominio. En lugar de partir de cero, se reutilizó una ontología llamada Generalized Upper Model (GUM), que ya había sido utilizada para varias lenguas como el inglés y el italiano. De esta forma, se validaba el modelo, mostrando su modularidad y reutilización. Para conseguir que sirviera para la generación en español se estudiaron detenidamente cada una de sus jerarquías, sus términos y sus categorías, y se realizaron las modificaciones pertinentes, tomando como base unos criterios establecidos al comienzo del proceso y que servían para garantizar su independencia de la aplicación y del dominio.

Claves: reutilización, generación de lenguaje natural, ontología.

1 Introducción.

Los sistemas de generación de lenguaje natural (GLN) requieren distintos tipos de información: sobre el dominio, el usuario, el contexto, el lexicon, la gramática, etc. En este artículo se abordan sólo aspectos relacionados con la información del dominio.

La fuente de información del dominio (base de datos, base de conocimientos, etc.) no es un componente del sistema de GLN como tal, pero es un recurso fundamental para él, ya que le proporciona lo que tiene que "decir".

Normalmente los datos del dominio son distintos a los términos que necesita un sistema de GLN. Esto implica la necesidad de determinar un mecanismo o una interfaz para pasar de los

términos del dominio (no lingüísticos) a los términos (lingüísticos) que puede usar el sistema de GLN.

Bateman [1996] menciona algunas de las posibles soluciones a este problema:

- Construir una codificación específica del dominio que refleje la forma en que se ha de expresar (lingüísticamente) la información.
- Crear tablas de correspondencias entre las categorías del dominio y los recursos lingüísticos del generador.
- Substituir o reemplazar progresivamente las configuraciones de los componentes de la representación del dominio por otras configuraciones más cercanas (en estructura o contenido) a la estructura de superficie que se quiere generar.
- Reducir la correspondencia entre los términos del dominio y los de GLN a una relación de subsunción, clasificando las categorías del dominio en términos de una jerarquía de objetos y relaciones generales, que se comportan de forma sistemática con respecto a sus posibles realizaciones lingüísticas. Este tipo de jerarquía la da, por ejemplo, la ontología *Generalized Upper Model* (GUM) [Bateman et al., 1995].

El caso que aquí se presenta está relacionado con un proyecto de GLN, al que a partir de ahora se hará referencia con el nombre de ONTOGENERATION¹ [Aguado et al., 1998], que consiste en la construcción de un sistema de consulta mediante la generación de textos en español en el dominio de las sustancias químicas. En este proyecto se optó por la última de las soluciones señaladas anteriormente, es decir, utilizar GUM.

En las siguientes secciones se describe brevemente la ontología, se explican las razones de esta elección, y se detalla el proceso que hubo que realizar para poder utilizarla posteriormente en el sistema de GLN. Por problemas de espacio, no se tratan aquí el uso de la ontología en el proceso de generación del sistema ni los aspectos de implementación.

2 Breve descripción de GUM

GUM es una ontología lingüística, lo que hace que esté ligada a la semántica de los componentes gramaticales de una lengua pero, a diferencia de otras como WordNet² [Miller, 1995], no describe la semántica de las palabras, sino que sus términos corresponden a la semántica que se puede expresar en unidades gramaticales mayores como grupos nominales, sintagmas verbales, sintagmas preposicionales, etc.

GUM está organizada en dos jerarquías: una de conceptos y una de relaciones. La jerarquía de conceptos representa las entidades semánticas básicas e incluye las configuraciones de procesos y las distintas clases de objetos y cualidades. La jerarquía de relaciones representa los

¹ Proyecto financiado por la UPM dentro del plan "Ayudas para la realización de proyectos de investigación y desarrollo, dirigidas a grupos potencialmente competitivos" (Referencia A9706).

² Aunque dentro del ámbito del procesamiento de lenguaje natural ambas se incluyen dentro de la categoría de ontologías lingüísticas, son de naturaleza muy distinta. WordNet (y sus derivados) se podrían considerar más bien una base de datos léxica.

participantes, las circunstancias involucradas en los procesos y las combinaciones lógicas entre ellos. La *Figura 1* muestra los primeros niveles de estas jerarquías.

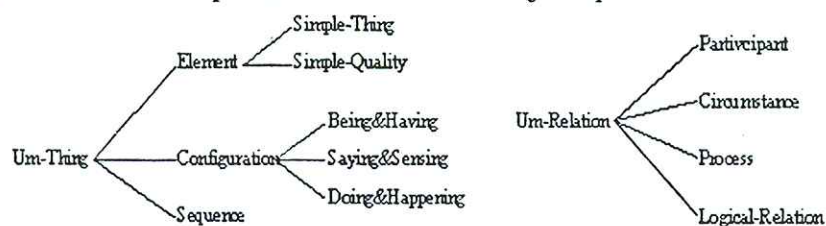


Figura 1: Primeros niveles de las jerarquías de GUM

Estas taxonomías tienen su origen en la gramática funcional de Halliday [1985], pero son aplicables en cualquier otra teoría.

3 Razones para elegir GUM.

Existen dos razones principales por las que se eligió GUM como base para desarrollar la interfaz entre el sistema de GLN de ONTOGENERATION y su fuente de información:

Primero, los trabajos previamente realizados con GUM mostraban que podía proporcionar una base sólida para la GLN cuando la organización del dominio estaba aislada de los detalles de su realización lingüística. Por tanto, el empleo de GUM como un nivel de interfaz aseguraba que:

1. No era necesario importar conocimientos lingüísticos a la fuente de información (la ontología *Chemicals* [Fernández, 1996]) para permitir la generación de textos, porque al clasificar los conceptos del dominio según la conceptualización de GUM, los términos del dominio heredaban de los de GUM las posibilidades para su expresión lingüística. De este modo no se interfería en la organización interna del dominio ni se violaba la modularidad deseada en un sistema.
2. Se simplificaba mucho la realización de una interfaz con el realizador gramatical³ del sistema (la herramienta KPML [Bateman, 1997]), puesto que gran parte de la información específica para el procesamiento del lenguaje estaba unida a la relación existente entre GUM y los recursos lingüísticos y, por tanto, no era necesario incluirla en las especificaciones de entrada que se pasaban al realizador.
3. Se reducía bastante la necesidad de reglas de procesamiento lingüístico específicas del dominio, pues GUM proporcionaba ese conocimiento mediante una organización conceptual, reutilizable, general e independiente del dominio y que podía emplearse para clasificar todo el conocimiento específico del dominio.

La segunda razón era la posibilidad de reutilización de GUM, demostrada en el trabajo que se había llevado a cabo para la GLN en varias lenguas. GUM proporcionaba un núcleo fijo lo

³ El realizador gramatical es el encargado de transformar una especificación de una frase en una oración gramaticalmente correcta. Dependiendo de la sofisticación de esas especificaciones (plantillas, estructuras sintácticas abstractas, etc.), el realizador gramatical tendrá que hacer una tarea más o menos compleja.

suficientemente general como para necesitar sólo variaciones pequeñas entre las lenguas. No era necesario adoptar una posición interlingüe, pero permitía minimizar los aspectos de la descripción semántica específicos de una lengua.

En el momento de comenzar el proyecto, la ontología GUM era válida para inglés, alemán e italiano. Se estudió la ontología para ver las modificaciones o adaptaciones que eran necesarias para la generación en castellano. El resultado de este proceso fue una nueva ontología, denominada GUME (GUM para español).

4 Criterios de diseño de GUME.

Los criterios para adaptar GUM al castellano se basaron en las pautas que se habían seguido durante el desarrollo de la ontología, desde la primera versión, creada para el sistema Penman [Bateman *et al.*, 1990], hasta GUM (criterios 1, 2, 3, 4, 5). También se añadieron otros criterios con el fin de mejorar la estructura y comprensión del modelo (criterios 6, 7, 8). Veamos cuáles eran esos criterios⁴.

1. Los conceptos y relaciones de la ontología eran necesarios para representar distintas expresiones léxico-gramaticales, y reflejar diferencias en su significado "experiencial"⁵.
2. Los tipos de especialización eran los mismos que en GUM: dimensión (si se describen distintos puntos de vista sobre una entidad), partición (si se representan alternativas mutuamente excluyentes que cubren todas las posibles especializaciones de una entidad), disyunción (si se representan alternativas mutuamente excluyentes que no cubren todas las posibles especializaciones de una entidad) y especialización simple (si no se puede decir nada sobre las propiedades disjuntas de los elementos que se van a introducir).
3. Las configuraciones con distinto número de participantes y circunstancias tenían distintas representaciones conceptuales. Estos participantes y circunstancias podían aparecer explícitamente o no.
4. La variación sintáctica en el orden de los argumentos sólo era relevante cuando la diferencia de patrones provocaba un cambio de significado. Obsérvense las oraciones:
 (A) *Cargó el vagón con heno.* (B) *Cargó heno en el vagón.*
 En (A) puede entenderse que cargó todo el vagón con esa mercancía, mientras que en (B) puede entenderse que, además de heno, puso otras mercancías en el vagón.
5. Se podía generar una misma expresión a partir de distintas categorías, según la perspectiva semántica que se tomase. Por ejemplo, la oración *Juan acabó el trabajo con Pedro* puede significar, 1) "Pedro" acompañó a "Juan" en la realización del trabajo (relación *Inclusive*); 2) "Juan" se sirvió de "Pedro" para realizarlo (relación *Generalized-Means*); 3) "Juan" fue el que acabó el trabajo que tenía que realizar con "Pedro" (relación *Attribute*).

⁴ El orden en el que aparecen enumerados los criterios no indica su orden de importancia.

⁵ Según Halliday, una oración tiene contribuciones de cuatro metafunciones, cada una de las cuales lleva consigo un conjunto de restricciones. La metafunción experiencial, que sirve para reflejar el conocimiento o la experiencia, es una de ellas.

6. Se intentaba mantener, hasta donde fuera posible, un estilo uniforme. Esto implicaba, por ejemplo, la modificación de las divisiones de algunas categorías de modo que no se mezclaran dimensiones con especializaciones simples.
7. Se procuraba que los cambios fueran mínimos, por dos razones: primero porque la ontología GUM (o alguna versión previa) ya estaba siendo utilizada y se había comprobado que funcionaba; y segundo, porque si las diferencias entre GUME y GUM eran pocas, sería más fácil su integración para conseguir una ontología que fuera válida para español, inglés, alemán e italiano. Si este criterio entraba en conflicto con el criterio 6; se optaba por lo más conveniente en cada situación.
8. Las nuevas categorías recibían un nombre en inglés, a efectos de uniformidad y coherencia terminológica.

5 Método para construir GUME.

En primer lugar, el hecho de partir de una ontología ya existente suponía que se había de seguir un método de construcción particular que sacara el máximo partido al trabajo que ya estaba realizado (en este caso, la ontología GUM). Se procedía del siguiente modo: para cada subjerarquía de GUM se estudiaba su descripción detalladamente. Después se buscaba un conjunto de comportamientos lingüísticos del castellano equiparables a los que representaba esa subjerarquía y se comparaban esos comportamientos con las explicaciones dadas en GUM. Si se observaban discrepancias, se proponían y realizaban las modificaciones correspondientes⁶, que habrían de estar de acuerdo con los criterios de diseño. En todo caso, se estudiaban posibles alternativas en la organización de la subjerarquía y si se encontraba una más adecuada, se realizaban las modificaciones pertinentes.

En segundo lugar, se disponía además de otro recurso: un corpus con ejemplos de las frases que se pretendía generar [Bernardos y Aguado, 2001]. La utilización del corpus aquí sirvió para dos fines fundamentales:

- Averiguar si era necesaria alguna nueva modificación. El procedimiento que se acaba de describir servía para adaptar GUM a partir de lo ya existente, pero podía ser necesario tener en cuenta alguna nueva perspectiva, propia del castellano que no se pudiera deducir a partir de la comparación indicada anteriormente. El modo de conseguir esto (dejando a un lado los momentos de inspiración) era utilizar textos de la lengua que se iba a generar. El corpus mencionado proporcionaba estos textos, aunque sólo para el dominio de los elementos químicos.
- Ayudar a validar la ontología. Es decir, confirmar que, al menos en el dominio de la fuente de información, GUME era correcta. Como se puede apreciar, este punto estaba en relación directa con el anterior. Puesto que sólo se utilizaban textos de un dominio concreto, no se

⁶ Normalmente estas diferencias residían en la división de la subjerarquía y en las restricciones que se imponían en la misma.

podía garantizar la validez del modelo en todos los dominios. Los criterios y procedimientos seguidos permitían suponer que si existían futuras modificaciones, éstas serían mínimas. Con el corpus se seguía el siguiente procedimiento. En cada texto se buscaban las categorías de GUM dentro de las que se podían clasificar los distintos componentes del mismo. Si existía alguna parte del patrón que no se podía clasificar dentro de GUM o que se podía clasificar sólo en una categoría muy abstracta, y esa expresión lingüística era general en español (no sólo propia del dominio de los elementos químicos), se creaba una subjerarquía que representara el significado de esa parte, teniendo siempre en cuenta los criterios señalados anteriormente.

6 Resultados.

Para ilustrar los puntos expuestos, se presenta un resumen cuantitativo del resultado obtenido tras el proceso de adaptación de GUM al castellano.

- Se identificaron 185 categorías de GUM válidas (119 conceptos y 66 relaciones).
- Se identificaron 6 categorías de GUM no válidas (3 conceptos y 3 relaciones).
- Se identificaron 20 categorías de GUM que necesitan modificaciones (12 conceptos y 8 relaciones) y se implementaron estas modificaciones.
- Se identificaron e implementaron 15 categorías nuevas (11 conceptos y 4 relaciones).
- La ontología GUME consta en total de 142 conceptos y 78 relaciones.

6.1 Ejemplo de creación de una categoría.

En GUM existe un concepto, denominado *Generalized-Positioning*, que se usa para representar situaciones donde se localiza un objeto en el espacio o en el tiempo. Algunos ejemplos de oraciones en inglés que se engloban en esta categoría son:

John is at the station (John está en la estación)

The concert is near here (El concierto es cerca de aquí)

Se puede apreciar que mientras en inglés se utiliza un único verbo para ubicar algo en un lugar, "to be", en español existen dos: "ser" y "estar". La utilización de uno u otro depende del elemento que se esté localizando. Así, en castellano, para localizar un suceso se utiliza el verbo "ser" y para localizar un objeto se utiliza el verbo "estar" [Seco, 1990].

Esta particularidad del español llevó a la introducción de dos nuevas categorías dentro de *Generalized-Positioning*, cuyos nombres son: *Generalized-Positioning-Event* (que representa la localización de un suceso) y *Generalized-Positioning-Object* (que representa la localización de un objeto). Y también supuso cambios en otra de las categorías: el concepto *Simple-Thing*, para distinguir entre un objeto y un suceso.

6.2 Ejemplo de eliminación de una categoría.

GUM tiene una categoría, *Name-of*, que recoge la relación que existe entre un nombre y quien lo tiene. Tras su estudio, no se apreció ninguna diferencia entre la posesión de cualquier objeto

(concepto *Ownership*) y la posesión de un nombre. Además, existe una categoría, llamada *Name-Relation*, que recoge expresiones del tipo *X is called Y* (*X se llama Y*), que son las que también corresponderían a esta categoría⁷.

Por otro lado tampoco parece que en inglés sea necesaria esta distinción⁸. Como se puede apreciar, el trabajo realizado también ha servido para depurar y, por tanto, mejorar GUM.

6.3 Ejemplo de modificación de una categoría.

La categoría *Generalized-Role-Relation* refleja la perspectiva generalizada sobre una entidad que desempeña un papel con respecto a otra. En GUM se indica que esta categoría puede representar las siguientes expresiones:

<domain> has <range> as <role> (*<dominio> tiene <ámbito> de <papel>*)

<domain> 's <role> is <range> (*el <papel> del <dominio> es <ámbito>*)

Por ejemplo:

John has Betty as a secretary (*John tiene a Betty de secretaria*)

Sin embargo, en una nota se señala que frases como *John's secretary is Betty* (*La secretaria de John es Betty*) tienen que ser generadas mediante el concepto *Identity*. Se puede observar que el patrón *<domain> 's <role> is <range>* se ajusta a esta oración. Su estudio llevó a la conclusión de que se trata de una identidad (correspondiente a *Identity*), ya que “la secretaria de John” y “Betty” son la misma persona, por lo que se desechó esa segunda opción⁹.

6.4 Ejemplo de categoría que queda igual.

Tanto en GUM como en GUME se considera la categoría *Existence* (Existencia), que corresponde a una configuración en la que sólo interviene una entidad. Mediante esta categoría se indica que hay o existe algo, como por ejemplo en *There is a block* (*Hay un libro*)

Como se puede ver, la forma de expresar el concepto de existencia es distinta en las dos lenguas. Esto no significa que se necesiten dos categorías separadas y que la de un modelo no valga en el otro, sino que las restricciones lingüísticas que se impongan a *Existence* serán diferentes en la gramática para el inglés y en la gramática para el castellano.

⁷ Esto no está en contradicción con el criterio de diseño número 2. No se elimina *Name-of* porque exista otra categoría que sirva para lo mismo, sino porque se considera que esta especialización no debe ser una *Generalized-Possession*. El hecho de que exista otra distinción (*Name-Relation*) con el mismo significado simplemente evita crear otra categoría que cubra su hueco.

⁸ Estos hechos se han comentado con el doctor Bateman, principal creador de la ontología, y serán tenidos en cuenta en futuras versiones de GUM.

⁹ Véase nota 8

6.5 Ejemplo de uso.

Como se indicaba en la introducción, el enlace necesario entre el conocimiento de un dominio concreto y GUM (o en este caso, GUME) se puede lograr mediante la clasificación de ese conocimiento según las categorías semánticas generales que proporciona GUM(E).

Una vez que una categoría del dominio (en ONTOGENERATION, la ontología Chemicals) se define para que sea más específica que una categoría de la taxonomía (véase la parte (a) de la Figura 2), hereda sus restricciones lingüísticas, por lo que ya se pueden hacer inferencias sobre cómo debe expresarse esa categoría del dominio. Por eso cuando un realizador como KPML recibe una especificación de oración como la de la parte (b) de la Figura 2, “sabe” que el concepto “fluorine” ha de tratarse como el concepto “object” de GUM(E)¹⁰.

<pre>(defconcept elements :is (:and penman-kb::object :primitive)) (defconcept halogen :is (:and elements :primitive)) (defconcept fluorine :is (:and halogen :primitive))</pre>	<pre>(CA / CLASS-ASCRPTION :DOMAIN (X / FLUORINE :DETERMINER THE) :RANGE (Y / HALOGEN :DETERMINER A))</pre>
(a) Parte de la unión de Chemicals con GUME	(b) Ejemplo de especificación de oración

Figura 2. Ejemplo para generar “El flúor es un halógeno.”.

Por otro lado, en la especificación de oración de la Figura 2 también se usan conceptos de GUM(E): la configuración “class-ascription” y los participantes “domain” y “range”. De este modo, no es necesario indicar al realizador qué función sintáctica desempeñan los elementos de la especificación. Además, esa configuración tiene unos valores por omisión por lo que tampoco se ha tenido que señalar el tiempo, el tipo de verbo, el orden de los argumentos, etc.

7 Conclusiones y trabajo futuro.

En esta comunicación se ha presentado la forma en la que se procedió en el proyecto ONTOGENERATION para enlazar la fuente de información del dominio con el sistema de GLN en castellano que se pretendía construir. Para ello se recurrió a una ontología lingüística llamada GUM, previamente utilizada en otros proyectos y con distintas lenguas, a la que se realizaron las modificaciones oportunas para poder reutilizarla en la generación de castellano.

Teniendo en cuenta los datos numéricos recogidos en la sección 6 cabe señalar que la ontología GUME es muy pequeña en comparación con otras ontologías utilizadas en PLN en castellano, como Eurowordnet¹¹ [Díez *et al.*, 1997]. Esto es debido a que, al igual que GUM, no contiene información sobre la semántica de las palabras, sino que la información se refiere a uni-

¹⁰ Normalmente también se establece una asociación entre algunos conceptos del dominio y los elementos léxicos que sirven para expresarlos por lo que tampoco es necesario incluir este dato en el plan de oración.

¹¹ Véase nota 2.

dades gramaticales mayores. Por eso, en el caso de ONTOGENERATION, además de crear un lexicon con términos propios de los elementos químicos, fue necesario introducir, por ejemplo, conocimiento sobre los artículos, los tiempos y las formas verbales, etc.

Sin embargo, el uso de GUME, a partir de la reutilización de GUM, presenta otras ventajas considerables. En primer lugar, se favorece la modularidad del sistema. Por un lado, se puede mantener separado el conocimiento general del dominio, de los detalles lingüísticos y, por otro, se simplifica la comunicación entre los distintos módulos, ya que parte de la información necesaria para generar los textos está ligada a la ontología y no es necesario pasarla de un módulo a otro. En segundo lugar, y relacionado con esto último, se puede afirmar que se reduce la complejidad de las tareas de generación, sobre todo en las elecciones sintácticas. Por último, si a esto se añade el uso de un entorno multilingüe como KPML, también se facilitará su utilización en un sistema multilingüe.

Los criterios establecidos y el procedimiento seguido para llevar a cabo la adaptación de GUM al español son independientes de la aplicación que se estaba construyendo y, por tanto, es posible concluir que se podrán aplicar para otras lenguas. Cabe destacar, asimismo, que el hecho de que durante el proceso de adaptación de GUM se utilicen ejemplos de los textos que se pretende que genere el sistema no implica que la ontología construida sea dependiente del dominio, puesto que estos textos sirven como ayuda pero no son los que guían el desarrollo de GUME, y el procedimiento seguido está pensado para garantizar que no se incluyan datos específicos del dominio. De cualquier modo, esto no se puede afirmar con total seguridad. Por ello, esta es una de las líneas sobre las que se está trabajando, la evaluación de GUME. El uso de este modelo en la generación de textos en el dominio de los elementos químicos ha demostrado su validez en este dominio. Sin embargo, será necesaria su aplicación en otros proyectos para averiguar si las categorías eliminadas, modificadas o introducidas son realmente correctas, y si son necesarios nuevos cambios.

Independientemente de los resultados obtenidos con la aplicación del modelo, existen varias categorías en GUME susceptibles de un análisis más profundo. Entre ellas pueden mencionarse las siguientes:

- La configuración *Doing&Happening* abarca un conjunto de expresiones muy amplio mientras que el número de especializaciones que contiene es muy pequeño. Así, expresiones tan distintas semánticamente como *La casa se derrumbó*, *Ha muerto* o *Surgió de repente* se clasifican en la misma categoría. Sería útil analizar la conveniencia de recoger estas diferencias mediante nuevas especializaciones dentro de *Doing&Happening*.
- El concepto *Sequence* todavía se encuentra en fase de estudio dentro del desarrollo de GUM y no se dispone de información sobre él. Esto hace que esta parte de la ontología sea la que más pueda evolucionar en un futuro. Por otro lado, su definición (situación compleja donde varias actividades están conectadas mediante alguna relación) hace pensar que está relacionada con la categoría *Logical-Relation*; por tanto, una revisión de los conceptos de la misma ayudará a discernir entre una categoría y otra, y servirá para especializar *Sequence*, y para definir otros tipos de relaciones lógicas.

Para finalizar, una de las tareas más interesantes es la integración de la ontología GUME con GUM, de modo que el modelo resultante pueda usarse en sistemas de generación multilingüe.

Referencias.

- [Aguado *et al.*, 1998]: G. Aguado, A. Bañón, J. Bateman, S. Bernardos, M. Fernández, A. Gómez, E. Nieto, A. Olalla, R. Plaza, A. Sánchez. "Ontogeneration: Reusing Domain and Linguistic Ontologies for Spanish Text Generation". Presentado en *Workshop on Applications of Ontologies and Problem Solving Methods, ECAI'98*, Brighton (Reino Unido), 1998.
- [Bateman *et al.*, 1990]: J. A. Bateman, R. T. Kasper, J. D. Moore and R. A. Whitney. "A General Organization of Knowledge for Natural Language Processing: the Penman Upper Model". *Technical Report*, USC/ISI, Marina del Rey, (EEUU), 1990.
- [Bateman *et al.*, 1995]: J. A. Bateman, R. Henschel and F. Rinaldi. "Generalized Upper Model 2.0: documentation", *Technical Report*, GMD/IPS, Darmstadt (Alemania), 1995.
- [Bateman, 1996]: J. Bateman, "Automated Discourse Generation". *Encyclopedia of Library and Information Science*, 1996.
- [Bateman, 1997]: J.A. Bateman. "Enabling technology for multilingual natural language generation: the KPML development environment". *Natural Language Engineering*, 1: 1-42. Cambridge University Press, Cambridge (Reino Unido), 1997.
- [Bernardos y Aguado, 2001]: S. Bernardos y G. Aguado. "A New Approach in Building a Corpus for Natural Language Generation Systems". *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics (CICLING-2001)*, México D. F. (México), febrero 2001.
- [Díez *et al.*, 1997]: J P.Díez, W. Peter y P. Vossen. "The Multilingual design of EuroWordNet". *Proceedings of ACL/EACL-97. Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid (España), julio 1997.
- [Fernández, 1996]: Fernández, M.. *Chemicals: Una Ontología de Elementos Químicos*, Trabajo Fin de Carrera, Facultad de Informática, Universidad Politécnica de Madrid (España), 1996.
- [Halliday, 1985]: M. A. K. Halliday. *An Introduction to Functional Grammar*. Edward Arnold, Londres (Reino Unido), 1985.
- [Miller, 1995]: G. A. Miller. "WordNet: a lexical database for English.". *Communications of the ACM* 38 (11): 39 – 41, 1995.
- [Seco, 1990]: Seco, R. *Manual de Gramática Española*, Aguilar, Madrid, 1990, 11^a ed.